“新基建”中台能力支撑智慧媒体发展

刘长明 卢 岚 徐 建
（北京北大方正电子有限公司，北京 100000）

摘　要：全国新基建产业按下快进键，移动化、数据化、智能化成为驱动行业进展的基础支撑，如何利用新一代信息技术手段，实现媒体行业自身基础能力的优化与升级，成为智慧媒体建设的新引擎和支撑？本文通过大数据与 AI 驱动的技术中台研究与实践，系统阐释如何在技术端发力，打造智慧媒体的“新基建”中台能力，并进一步支撑智慧媒体的发展。

关键词：大数据；人工智能；技术中台；智媒体；新基建；“数据 +AI”双中台　**中图分类号：**TP246　**文献标识码：**A
本文著录格式：刘长明，卢岚，徐建 .“新基建”中台能力支撑智慧媒体发展 [J]. 中国传媒科技，2021（02）：10−13.

1. 媒体为什么要构建技术中台

以 5G 为代表的通信网络基础设施，以数据中心、智能计算中心为代表的算力基础设施，和以人工智能、云计算、区块链等为代表的新技术基础设施是全国新基建背景下信息与融合基础设施建设的三大核心技术。新技术支撑体系下的媒体形态将充分融合，而支撑智慧媒体的能力建设也面临新的技术架构发展——技术中台。

技术中台的目标并不是为了解决单点问题，而是提供可复用的“基础服务和能力”的整合平台 [2]，大数据和人工智能技术都是基础性的“资源、服务和能力”，是新的“系统架构思路”和“系统建设模式”。大数据和 AI 不仅为内容生产、传播、经营与业务拓展提供基础性服务和支撑，更可以拓展到数据服务、智能交互、媒体智库、智慧政务、智慧城市等新业态。因此中台战略是大数据和 AI 能力的最佳构建方式，通过基础数据、服务与能力的整合，为上层业务赋能，同时推动媒体整体技术架构的转型升级。中台能力建设，可以为媒体持续快速新业务的发展，提供长期的支撑系统。

方正采用“数据 +AI”双中台架构，利用当前快速布局的 5G 网络，为数据传输与创新产业的发展提供技术支撑。依托大数据提供海量数据资源与分析处理，为生产服务提供有效信息；通过 AI 人工智能应用提供更为成熟、科学的决策方案并进行深度信息处理，为上层智慧媒体业务提供支撑，构成智慧媒体新基建的生态结构，赋能数字经济时代的高质量发展（图 1）。



图 1　方正基于双中台的超融合 4.0 解决方案

2. 怎样构建技术中台

2.1　数据中台

数据中台是聚合和治理跨域数据，为智能经济的底层基础，是产业数字化转型的必然要求。[3]

经过几十年的技术积累，中国媒体行业大部分已经完成多轮信息化建设过程，构建起来自不同技术供应商、不同数据存储标准的内部数据资产管理体系。同时，随着智慧媒体建设进程的推进，需要导入海量互联网内容数据和用户数据，以及引入三方智库报告、知识库等多种数据形式。客观需求上，传统的架构方式数据的整合、存储和应用越发凸显其弊端，数据中台将数据抽象封装成服务，通过 API 的方式提供数据服务，Data API 是数据中台的核心，它是连接前台和后台的桥梁。通过引入“数据中台”，采用分布式、组件化技术体系，对海量异构数据进行智能化采集、计算、存储、加工，形成大数据资产层，进而为客户及拓展业务提供高效数据服务。方正利用在媒体行业深耕多年的业务经验、技术经验及项目实施经验等方面的持续积累，形成得天独厚的数据管理优势，通过对多源数据接入管理、数据质量控制、数据组织存储和对外多维服务输出，实现从数据到数据服务的能力（图 2）。

2.1.1　数据采集和接入服务

数据中台具备标准化、跨体系、跨模态的开放服务体系。[3] 适配各业务系统接入，便于数据交换共享，适应需求变更及扩展，可根据需求快速构建数据模型。通过丰富的采集工具集，支持多源数据接入，实现各系统数据的实时获取，并通过插件化的方式，实现对接入数据的清洗、智能排重、转换及标准化、完整性唯一性校验等处理，保证接入数据的完整性和一致性。并



图 2　开放的数据中台架构图

对接入数据通过智能分析技术进行内容挖掘，实现对文本、图像、视频等非结构化数据的结构化自动标签。

2.1.2　*数据质量管理*

其核心是建立统一的数据标准体系，对数据的抓取、清洗、入库进行全方位的实时跟踪与质检，通过系统后台的数据监控平台对入库数据进行检查，支持从系统中生成数据状态报告。当数据处理过程中出现突发异常状况时，显示故障状况，并以多种报警提示方式向预先设置好的多个指定终端发送预警报告。

2.1.3　*数据资产管理*

数据资产管理是规划和组织管理数据，必须围绕媒体业务的特征和需求痛点来确定数据资产规划、数据价值确认，采用分布式架构可以随着业务发展规模而灵活扩充，对自有数据资源和接收的互联网数据资源进行持久化组织存储、资源确权，逐步形成大数据资产中心。

数据资产根据其来源与属性特征分三大类，包括由生产数据、音视频数据和新媒体成品构成的自有内容数据资产；由互联网稿件素材、互联网价值信息、互联网聚合专题信息构成的互联网数据资产；由各业务用户资源整合，用户行为、用户画像构成的用户数据资产。不同类别采用适合的数据资产组织方式，最大化保留其数据特征信息。

2.1.4　*数据挖掘分析*

数据挖掘分析实现大规模数据的挖掘分析和数据模型构建，通过与 AI 智能能力的深度结合，以机器学习、深度学习等算法模型为基础，针对媒体应用场景，综合运用多种算法分析结果进行有效整合，进一步结合媒体领域的业务知识，形成多种应用模型。例如热点模型、传播模型、事件模型、影响力模型，从海量的数据中提取知识，挖掘出潜在的、有价值的内容信息。数据应用模型采用插件式管理模式，可以根据业务需要随时增减，也为未来业务发展预留空间。

2.1.5　*数据服务输出*

数据中台的核心目标是为各类应用提供统一、开放、类型丰富的数据服务。数据服务层为各应用系统提供标准数据访问接口，数据服务主要包括数据统一检索、数据推送、数据查看和选用，对外提供标准数据接口 API，支持主题式数据服务开发。

综上，数据中台是构建上层智慧媒体百花齐放式业务形态的基础，数据服务层极大降低业务层对数据的专一捆绑与束缚，从而更低成本、高效率地孵化业务创新。数据中台所践行的是一种数据完全开放的理念，尽其所能消除各系统之间的数据壁垒，构建数据流畅通无阻的生态圈。

2.2　AI 智能中台

智慧媒体构建中需要引入大量 AI 智能能力，整个媒体行业信息化发展已经走到了智能化能力运用的快速增长期，通过智能化有效提升媒体的运营效率。但是，有一个障碍一直在困扰大家，要想让 AI 能力发挥得更充分，就需要其与业务系统的融入更紧密、更深入，而这也意味随着接入服务的多元化，更换 AI 能力的成本高、周期长。当前处在 AI 智能化发展的爆发期，多种智能工具的接入、试错、灵活选择和更换成为必需，AI 中台正是在这样的背景下诞生的（图 3）。



图 3　AI 智能中台架构图

AI 中台提供开放的基于微服务组件的智能服务市场，以“构建开放的 AI 能力整合中台，为媒体业务赋能”为目标，支持智能组件的快速接入和输出。依托方正多年的技术积累及与北京大学等国内知名高校和企业的多方面合作，提供包括自然语言处理、计算机视觉、知识图谱、区块链等智能技术能力，形成近 60 个智能组件。构建了一套标准的 AI 能力管理和接入体系，形成了一系列面向媒体应用的智能组件，并通过 IDS 实现对接口的细粒度高效共享和访问控制，采用微服务架构满足智能能力的持续扩大和拓展，为融媒体生产系统、媒体大数据系统等业务提供智能能力支撑，也可以提供媒体单位第三方系统所需的智能能力输出。AI 中台作为媒体行业的智能化技术支撑，深度浸入媒体业务场景，提供辅助媒体全流程的智能化场景能力，全面助力媒体智能化建设。即：通过 AI 中台与智慧媒体的业务进行深入融合，所有接入的 AI 智能能力，都可以被其转化为标准服务层，在 AI 算法能力层不断变化的同时，保持与业务层的衔接

是持续、稳定的状态。

2.2.1 开放的 AI 能力平台

以 PaaS 平台为基础，采用深度学习和微服务技术架构，构建 AI 组件库。基于文本、图片、视音频相关的智能处理技术及多源 AI 服务能力的集成与管理，以标准接口方式向应用系统输出 AI 智能处理能力。

通过自然语言处理技术实现对文本数据的自动分类与生产辅助，包括关键词提取、实体抽取、实体关系抽取、情感分析、文本分类、自动摘要、实体链接与消歧等；通过语音处理技术，实现对语音的识别、合成与视音频文件的转写等；通过图像处理技术，实现面向图像的文字识别、相似度比对、人脸识别、图片标签、图像检索等；通过视频处理技术，实现视频中镜头分割与关键帧提取、视频标签、视频检索等。综合利用多模态内容分析与识别、语义分析、VR/AR、虚拟主播、机器创作、知识图谱、区块链版权等新技术集群的研发和渗透，为媒体内容生产与呈现开拓想象空间。

2.2.2 基于场景的智能应用和服务

深度结合应用场景，围绕生产多样化、审核深入化、传播互动化等核心业务需求，对文本、语音、图片、视频多模态数据进行综合分析，提供智能策划、智能写作、智能生产辅助、智能标签、智能风控、智能分发与传播、智能推荐等基于场景应用的智能能力组件。

针对内容生产，利用自然语言处理、语音识别与计算机视觉技术，基于多维度对内容进行结构化标签提取，包含地域、领域、情感、实体、关键词、场景等，可广泛应用于内容的管理、搜索和推荐，用于快速获取业务中所需要的多模态数据。

针对内容安全问题，采用拼写纠错、语法纠错、预训练、目标检测等多种深度学习技术，辅助媒体领域知识库，使内容审核能快速有效的抓取到不合规的信息，保障传播内容的真实性与合法性，如：错别字、敏感词、涉黄、涉暴涉恐、政治敏感、微商广告等相关信息，大幅节省内容审核的人力成本，提高业务处理的准确性和高效性，为内容发布保驾护航。

针对图文、音视频生产，提供图像主体检测、风格转换、图文一键转视频、视频转 gif 动画、视频封面选取等便捷的智能工具。

针对信息传播，对内容的情绪、新闻价值等因素进行综合分析，对信息和稿件进行预判，针对新闻阅读量、点赞量、评论量、转发量的影响因素，综合评估得到新闻的传播力指数，预测稿件影响力，为媒体业务修正或转型提供支撑和依据。

3. 数据 &AI 双中台的应用场景和建设价值

对于什么应用场景下需要使用中台相关技术，我们基于对中台的应用与实践，来谈谈其中的体会。

3.1 双中台为媒体数据资产建设提供有力支撑

如今媒体的产品形态多元化，各渠道发布的都是经过媒体加工生产的成品数据或半私有数据，都是媒体的数字资产，都需要收集并存储管理起来，伴随现在的技术，大量互联网数据，分析数据、合作数据，都可以实现分析、挖掘，为数据应用提供支撑。而传统数据管理方式在新数据接入和管理上适应力弱，耗费巨大项目实施成本，拓展性也不够。

引入数据中台，媒体在数据资产建设中形成完整、开放的数据管理，通过数据的规划和建设，让集团各媒体单位的数据变成统一的标准，将所有可能用上的、所有对集团有可能有价值的数据都系统整理出来， 数据中台是跨域的，能够将不同组织、不同类型的数据整合并提供服务的开放数据平台，让集团的人都知道数据资产在哪里，在数据安全的基础上，为数据相关者、价值创造者提供服务。例如方便地实现各种类型数据的混合检索，可以方便检索寻找需要的稿件，通过智能化分析可以让编辑快速看到稿件的摘要，关键词，以及网络的反馈效果，为编辑生产引用进行判断，对选中的单篇及批量稿件可一键选用至生产系统，为编辑高效生产提供帮助。

数据中台为集团提供了强大的数据资产的获取和存储的能力。为日益增长的数据资源在鉴别、选择、管理上提供支撑。

3.2 双中台将数据和智能实现标准化的能力输出

过去的数据系统，是前后台一体，造成数据结构与应用业务高度融合，应用变化则数据结构变化，束缚了数据价值的挖掘。通过中台技术，将前后端分离，数据被封装为数据服务与数据能力，AI 计算池被封装为 AI 智能能力，而业务应用不再关注底层的数据结构，通过微组件的方式，直接去调取所需的数据能力和 AI 智能能力，简单的输入与输出，即可达到业务目标。

同样的理念，数据的分析与挖掘也采用插件方式来构建，通过这样的模型抽象，所汇聚的内部生产数据、互联网数据及用户数据进行相互整合，并由此产生新的数据关联，形成新的数据价值或新的知识节点。例如，新媒体成品数据的存储与传播、运营分析的价值数据相结合，不仅让数据作为历史被全方位、多维度地保留下来，更是为当前、未来，媒体单位如何更有效地开展新媒体运营业务提供数据分析依据，从宏观上指导主管领导去决策，从微观上优化执行层面的日常工作。

双中台根据业务场景的需求进行组件开发，提供更好的数据能力或智能能力输出。例如智能微服务组件，包括组件体系、标准接口规范、CICD 规范，具有垂直开发横向扩展及快速复用等特点。采用敏捷开发模式，自由选择合适的技术快速迭代部署，根据服务需求选择合适的虚拟机并匹配服务资源要求的硬件，应用程序分解成可管理的模块和服务，单个服务可实现快速开发、简单维护。

3.3　双中台是一个生态平台可激发更多的业务创新

双中台的应用实践也依据具体项目是可弹性扩展的，在简单如媒体资源库的项目中，数据接入服务可以有效解决异构数据的整合与综合利用；AI 智能分析能力可以提升资源，尤其是图片、视音频资源的标签化水平和被精准检索到的概率。而对于一个完整的数据资产建设项目，技术双中台则更利于系统整体架构的搭建，充分支撑当前业务需求，且可以随着业务发展和孵化的数据服务，引入更多先进的智能能力，从而推动业务的创新。

双中台让数据服务和智能能力不再与业务应用紧密捆绑，具有充分的开放性，其标准化可以让更多服务接入或者让更多应用去方便调用服务，使应用场景的满足与服务能力的提升各自去发展，相互不约束。使业务更换服务能力的成本更低，使服务能力能灵活组合、去支撑更多的创新业务。

中台是一个生态平台，在数据中台上面会不断生长各种数据服务，数据服务需求可以被记录、可被跟踪、可被监控。不断将所挖掘的数据价值与数据能力，源源不断地输入到整个智慧媒体的业务中，从而践行从数据到数据服务能力的理念。伴随智能能力变化，AI 中台可以不断接入和输出最新优质的智能能力，供业务系统灵活选择。数据和 AI 的结合为智慧媒体发展提供更多新的可能。

4. 展望

媒体即将迎来“信息随心至，万物皆可及”的智媒时代。瞄准新技术，享受大数据与人工智能的服务，是新时代智慧媒体建设的最优路径。“数据 +AI”双中台技术，以安全可靠、稳健成熟的实用主义为指导，作为媒体“新基建”为智慧媒体发展提供有力支撑并为业务创新提供土壤，使媒体角色从“资讯内容提供者”向“城市信息服务者”转变。媒

参考文献

[1] 方正电子 . 智慧媒体大家说 | 新基建背景下，媒体如何用好大数据和 AI？ [EB/OL].2020-08-10.

[2] 郑伟 . 媒体融合的“数”与“智”[J]. 传媒评论，2020（10）：16-17.

[3] 卢岚 . 基于“数据 +AI”双中台技术打造智慧媒体新基建 [C].2020 年中国报业技术年会论文集，2020（11）：650-655.

作者简介：刘长明（1965-），男，黑龙江哈尔滨市，北京北大方正电子有限公司副总裁；卢岚（1976-），女，青海，北京北大方正电子有限公司大数据产品部总经理；徐建（1976-），男，陕西，北京北大方正电子有限公司产品总监。

（责任编辑：陈旭管）

新闻行业奖项一览表（部分）

为激励新闻工作者提升业务能力，研发新技术、创新新应用，推动我国新闻行业发展与进步，特梳理部分有影响力的新闻行业奖项，为大家申报奖项提供一些参考。

根据全国评比达标表彰工作协调小组办公室 2015 年公布的《全国评比达标表彰保留项目目录》，及 2019 年 12 月 2 日至 2020 年 12 月 9 日中华人民共和国人力资源和社会保障部官网公布的《关于评比达标表彰项目的公示》等公开信息，梳理出部分新闻行业奖项。

其中，党群系统等中央单位评比达标表彰保留项目（127 项）中，媒体可参评的奖项有（注：括号内为主办单位）：

1. 党风廉政建设好新闻奖（中央纪委监察部）；
2. 全国政法优秀新闻作品（中央政法委）（原为：全国政法综治优秀新闻作品，主办部门：中央综治办，2019 年 12 月更名）；
3.“中华环保世纪行”宣传活动新闻奖（全国人大环资委）；
4. 人大新闻奖（全国人大常委会办公厅）；
5. 全国政协好新闻（全国政协办公厅）；
6. 中国新闻奖（中国记协）；
7. 长江韬奋奖（中国记协）
8. 残疾人事业好新闻奖（中国残联）；

中央国家机关评比达标表彰保留项目（381 项）中，媒体可参评的奖项有：
1. 全国广播电视系统先进集体、劳动模范和先进工作者（原新闻出版广电总局，现为国家新闻出版署）；
2. 全国新闻出版系统先进集体、先进工作者和劳动模范（原新闻出版广电总局，现为国家新闻出版署）；
3. 中国出版政府奖（原新闻出版广电总局，现为国家新闻出版署）；
4. 版权保护示范单位（原新闻出版广电总局，现为国家新闻出版署）；
5. 中华图书特殊贡献奖（原新闻出版广电总局，现为国家新闻出版署）；
6. 王选新闻科学技术奖（中国新闻技术工作者联合会）；
7. 中华优秀出版物奖（中国出版工作者协会）；
8. 韬奋出版奖（中国出版工作者协会）；
9. 中国经济新闻（大赛）暨经济新闻人物评选（原中国经济报刊协会，现更名为中国经济传媒协会）；
10. 中国广播电视大奖（国家广播电视总局，2020 年 6 月新增）。

其中王选新闻科学技术奖是经国家科技奖励办公室批准设立的、新闻界跨媒体的科技奖项，是以推动新闻传媒界的科技进步，服务于广大相关从业人员为宗旨。“王选奖”的项目奖和人才奖分别每两年评审一次，对激励我国广大新闻科技工作者，推动传媒科技创新和发展，服务于新闻报道和传媒的战略转型，起到了积极的鼓励和导向作用。

关于“王选奖”更多评奖信息，请登录中国新闻技术工作者联合会官网（http：//www.capt.cn/）查询。

注：2018 年机构改革后，主办单位名称有的发生了变化。如有疏漏之处，敬请大家斧正。